

# KILLTEST

20% de descuento en todos los productos

Preguntas del examen

DP-700

Implementing Data  
Engineering Solutions  
Using Microsoft Fabric



---

## 1. Topic 1, Contoso, Ltd

### Case Study

#### Overview

This is a case study. Case studies are not timed separately. You can use as much exam time as you would like to complete each case. However, there may be additional case studies and sections on this exam. You must manage your time to ensure that you are able to complete all questions included on this exam in the time provided.

To answer the questions included in a case study, you will need to reference information that is provided in the case study. Case studies might contain exhibits and other resources that provide more information about the scenario that is described in the case study. Each question is independent of the other questions in this case study.

At the end of this case study, a review screen will appear. This screen allows you to review your answers and to make changes before you move to the next section of the exam. After you begin a new section, you cannot return to this section.

#### To start the case study

To display the first question in this case study, click the Next button. Use the buttons in the left pane to explore the content of the case study before you answer the questions. Clicking these buttons displays information such as business requirements, existing environment, and problem statements. If the case study has an All Information tab, note that the information displayed is identical to the information displayed on the subsequent tabs. When you are ready to answer a question, click the Question button to return to the question.

#### Overview. Company Overview

Contoso, Ltd. is an online retail company that wants to modernize its analytics platform by moving to Fabric. The company plans to begin using Fabric for marketing analytics.

#### Overview. IT Structure

The company's IT department has a team of data analysts and a team of data engineers that use analytics systems.

The data engineers perform the ingestion, transformation, and loading of data. They prefer to use Python or SQL to transform the data.

The data analysts query data and create semantic models and reports. They are qualified to write queries in Power Query and T-SQL.

#### Existing Environment. Fabric

---

Contoso has an F64 capacity named Cap1. All Fabric users are allowed to create items.

Contoso has two workspaces named WorkspaceA and WorkspaceB that currently use Pro license mode.

#### Existing Environment. Source Systems

Contoso has a point of sale (POS) system named POS1 that uses an instance of SQL Server on Azure Virtual Machines in the same Microsoft Entra tenant as Fabric. The host virtual machine is on a private virtual network that has public access blocked. POS1 contains all the sales transactions that were processed on the company's website.

The company has a software as a service (SaaS) online marketing app named MAR1. MAR1 has seven entities. The entities contain data that relates to email open rates and interaction rates, as well as website interactions. The data can be exported from MAR1 by calling REST APIs. Each entity has a different endpoint.

Contoso has been using MAR1 for one year. Data from prior years is stored in Parquet files in an Amazon Simple Storage Service (Amazon S3) bucket. There are 12 files that range in size from 300 MB to 900 MB and relate to email interactions.

#### Existing Environment. Product Data

POS1 contains a product list and related data.

The data comes from the following three tables:

- Products
- ProductCategories
- ProductSubcategories

In the data, products are related to product subcategories, and subcategories are related to product categories.

#### Existing Environment. Azure

Contoso has a Microsoft Entra tenant that has the following mail-enabled security groups:

- DataAnalysts: Contains the data analysts
- DataEngineers: Contains the data engineers

Contoso has an Azure subscription.

The company has an existing Azure DevOps organization and creates a new project for repositories that relate to Fabric.

#### Existing Environment. User Problems

The VP of marketing at Contoso requires analysis on the effectiveness of different types of email content. It typically takes a week to manually compile and analyze the data. Contoso wants to reduce the time to less than one day by using Fabric.

The data engineering team has successfully exported data from MAR1. The team experiences transient connectivity errors, which causes the data exports to fail.

---

## Requirements. Planned Changes

Contoso plans to create the following two lakehouses:

- Lakehouse1: Will store both raw and cleansed data from the sources
- Lakehouse2: Will serve data in a dimensional model to users for analytical queries

Additional items will be added to facilitate data ingestion and transformation.

Contoso plans to use Azure Repos for source control in Fabric.

## Requirements. Technical Requirements

The new lakehouses must follow a medallion architecture by using the following three layers: bronze, silver, and gold. There will be extensive data cleansing required to populate the MAR1 data in the silver layer, including deduplication, the handling of missing values, and the standardizing of capitalization.

Each layer must be fully populated before moving on to the next layer. If any step in populating the lakehouses fails, an email must be sent to the data engineers.

Data imports must run simultaneously, when possible.

The use of email data from the Amazon S3 bucket must meet the following requirements:

- Minimize egress costs associated with cross-cloud data access.
- Prevent saving a copy of the raw data in the lakehouses.

Items that relate to data ingestion must meet the following requirements:

- The items must be source controlled alongside other workspace items.
- Ingested data must land in the bronze layer of Lakehouse1 in the Delta format.
- No changes other than changes to the file formats must be implemented before the data lands in the bronze layer.
- Development effort must be minimized and a built-in connection must be used to import the source data.
- In the event of a connectivity error, the ingestion processes must attempt the connection again.

Lakehouses, data pipelines, and notebooks must be stored in WorkspaceA. Semantic models, reports, and dataflows must be stored in WorkspaceB.

Once a week, old files that are no longer referenced by a Delta table log must be removed.

## Requirements. Data Transformation

In the POS1 product data, ProductID values are unique. The product dimension in the gold layer must include only active products from product list. Active products are identified by an IsActive value of 1.

Some product categories and subcategories are NOT assigned to any product. They are NOT analytically relevant and must be omitted from the product dimension in the

---

gold layer.

### Requirements. Data Security

Security in Fabric must meet the following requirements:

- The data engineers must have read and write access to all the lakehouses, including the underlying files.
- The data analysts must only have read access to the Delta tables in the gold layer.
- The data analysts must NOT have access to the data in the bronze and silver layers.
- The data engineers must be able to commit changes to source control in WorkspaceA.

You need to ensure that the data analysts can access the gold layer lakehouse.

What should you do?

- A. Add the DataAnalyst group to the Viewer role for WorkspaceA.
- B. Share the lakehouse with the DataAnalysts group and grant the Build reports on the default semantic model permission.
- C. Share the lakehouse with the DataAnalysts group and grant the Read all SQL Endpoint data permission.
- D. Share the lakehouse with the DataAnalysts group and grant the Read all Apache Spark permission.

Answer: C

Explanation:

Data Analysts' Access Requirements must only have read access to the Delta tables in the gold layer and not have access to the bronze and silver layers.

The gold layer data is typically queried via SQL Endpoints. Granting the Read all SQL Endpoint data

permission allows data analysts to query the data using familiar SQL-based tools while restricting access to the underlying files.

## 2.HOTSPOT

You need to recommend a method to populate the POS1 data to the lakehouse medallion layers.

What should you recommend for each layer? To answer, select the appropriate options in the answer area. NOTE: Each correct selection is worth one point.

Bronze layer:

- A Dataflow Gen2 dataflow
- A notebook
- A pipeline Copy activity
- A pipeline stored procedure

Silver layer:

- A Dataflow Gen2 dataflow
- A notebook
- A pipeline Copy activity
- A pipeline stored procedure

Answer:

Bronze layer:

- A Dataflow Gen2 dataflow
- A notebook
- A pipeline Copy activity
- A pipeline stored procedure

Silver layer:

- A Dataflow Gen2 dataflow
- A notebook
- A pipeline Copy activity
- A pipeline stored procedure

Explanation:

---

### Bronze Layer: A pipeline Copy activity

The bronze layer is used to store raw, unprocessed data. The requirements specify that no transformations should be applied before landing the data in this layer. Using a pipeline Copy activity ensures minimal development effort, built-in connectors, and the ability to ingest the data directly into the Delta format in the bronze layer.

### Silver Layer: A notebook

The silver layer involves extensive data cleansing (deduplication, handling missing values, and standardizing capitalization). A notebook provides the flexibility to implement complex transformations and is well-suited for this task.

3.You need to ensure that usage of the data in the Amazon S3 bucket meets the technical requirements.

What should you do?

- A. Create a workspace identity and enable high concurrency for the notebooks.
- B. Create a shortcut and ensure that caching is disabled for the workspace.
- C. Create a workspace identity and use the identity in a data pipeline.
- D. Create a shortcut and ensure that caching is enabled for the workspace.

Answer: B

Explanation:

To ensure that the usage of the data in the Amazon S3 bucket meets the technical requirements, we must address two key points:

- Minimize egress costs associated with cross-cloud data access: Using a shortcut ensures that Fabric does not replicate the data from the S3 bucket into the lakehouse but rather provides direct access to the data in its original location. This minimizes cross-cloud data transfer and avoids additional egress costs.
- Prevent saving a copy of the raw data in the lakehouses: Disabling caching ensures that the raw data is not copied or persisted in the Fabric workspace. The data is accessed on-demand directly from the Amazon S3 bucket.

### 4.HOTSPOT

You need to create the product dimension.

How should you complete the Apache Spark SQL code? To answer, select the appropriate options in the answer area. NOTE: Each correct selection is worth one point.

SELECT ProductID, ProductNumber, ProductName, ModelName, SubCategoryName, CategoryName

FROM ContosoLake.Products p

FULL JOIN
INNER JOIN
LEFT ANTI JOIN
LEFT OUTER JOIN
OUTER JOIN

ContosoLake.ProductSubCategories s ON p.SubCategoryID = s.SubCategoryID

FULL JOIN
INNER JOIN
LEFT ANTI JOIN
LEFT OUTER JOIN
OUTER JOIN

ContosoLake.ProductCategories c ON c.CategoryID = s.CategoryID

WHERE

CategoryID = 1;
CategoryName is not null;
IsActive = 1;
IsActive is not null;
ProductNumber is not null;
SubCategoryID = 1;
SubCategoryName is not null;

Answer:

SELECT ProductID, ProductNumber, ProductName, ModelName, SubCategoryName, CategoryName

FROM ContosoLake.Products p

FULL JOIN
INNER JOIN
LEFT ANTI JOIN
LEFT OUTER JOIN
OUTER JOIN

ContosoLake.ProductSubCategories s ON p.SubCategoryID = s.SubCategoryID

FULL JOIN
INNER JOIN
LEFT ANTI JOIN
LEFT OUTER JOIN
OUTER JOIN

ContosoLake.ProductCategories c ON c.CategoryID = s.CategoryID

WHERE

CategoryID = 1;
CategoryName is not null;
IsActive = 1;
IsActive is not null;
ProductNumber is not null;
SubCategoryID = 1;
SubCategoryName is not null;

Explanation:

Join between Products and ProductSubCategories:



---

- Use an INNER JOIN.

- The goal is to include only products that are assigned to a subcategory. An INNER JOIN ensures that only matching records (i.e., products with a valid subcategory) are included.

Join between ProductSubCategories and ProductCategories:

- Use an INNER JOIN.

- Similar to the above logic, we want to include only subcategories assigned to a valid product category. An INNER JOIN ensures this condition is met.

WHERE Clause

Condition: IsActive = 1

Only active products (where IsActive equals 1) should be included in the gold layer.

This filters out inactive products.

5. You need to populate the MAR1 data in the bronze layer.

Which two types of activities should you include in the pipeline? Each correct answer presents part of the solution. NOTE: Each correct selection is worth one point.

A. ForEach

B. Copy data

C. WebHook

D. Stored procedure

Answer: AB

Explanation:

MAR1 has seven entities, each accessible via a different API endpoint. A ForEach activity is required to iterate over these endpoints to fetch data from each one. It enables dynamic execution of API calls for each entity.

The Copy data activity is the primary mechanism to extract data from REST APIs and load it into the bronze layer in Delta format. It supports native connectors for REST APIs and Delta, minimizing development effort.

6. Topic 2, Litware, Inc

Case Study

Overview

This is a case study. Case studies are not timed separately. You can use as much exam time as you would like to complete each case. However, there may be additional case studies and sections on this exam. You must manage your time to ensure that you are able to complete all questions included on this exam in the time provided.

To answer the questions included in a case study, you will need to reference information that is provided in the case study. Case studies might contain exhibits and

---

other resources that provide more information about the scenario that is described in the case study. Each question is independent of the other questions in this case study.

At the end of this case study, a review screen will appear. This screen allows you to review your answers and to make changes before you move to the next section of the exam. After you begin a new section, you cannot return to this section.

To start the case study

To display the first question in this case study, click the Next button. Use the buttons in the left pane to explore the content of the case study before you answer the questions. Clicking these buttons displays information such as business requirements, existing environment, and problem statements. If the case study has an All Information tab, note that the information displayed is identical to the information displayed on the subsequent tabs. When you are ready to answer a question, click the Question button to return to the question.

### Overview

Litware, Inc. is a publishing company that has an online bookstore and several retail bookstores worldwide. Litware also manages an online advertising business for the authors it represents.

### Existing Environment. Fabric Environment

Litware has a Fabric workspace named Workspace1. High concurrency is enabled for Workspace1.

The company has a data engineering team that uses Python for data processing.

### Existing Environment. Data Processing

The retail bookstores send sales data at the end of each business day, while the online bookstore constantly provides logs and sales data to a central enterprise resource planning (ERP) system.

Litware implements a medallion architecture by using the following three layers: bronze, silver, and gold. The sales data is ingested from the ERP system as Parquet files that land in the Files folder in a lakehouse. Notebooks are used to transform the files in a Delta table for the bronze and silver layers. The gold layer is in a warehouse that has V-Order disabled.

Litware has image files of book covers in Azure Blob Storage. The files are loaded into the Files folder.

### Existing Environment. Sales Data

Month-end sales data is processed on the first calendar day of each month. Data that is older than one month never changes.

In the source system, the sales data refreshes every six hours starting at midnight

---

each day.

The sales data is captured in a Dataflow Gen1 dataflow. When the dataflow runs, new and historical data is captured.

The dataflow captures the following fields of the source:

- Sales Date
- Author
- Price
- Units
- SKU

A table named AuthorSales stores the sales data that relates to each author. The table contains a column named AuthorEmail. Authors authenticate to a guest Fabric tenant by using their email address.

Existing Environment. Security Groups

Litware has the following security groups:

- Sales
- Fabric Admins
- Streaming Admins

Existing Environment. Performance Issues

Business users perform ad-hoc queries against the warehouse. The business users indicate that reports against the warehouse sometimes run for two hours and fail to load as expected. Upon further investigation, the data engineering team receives the following error message when the reports fail to load: "The SQL query failed while running."

The data engineering team wants to debug the issue and find queries that cause more than one failure.

When the authors have new book releases, there is often an increase in sales activity. This increase slows the data ingestion process.

The company's sales team reports that during the last month, the sales data has NOT been up-to-date when they arrive at work in the morning.

Requirements. Planned Changes

Litware recently signed a contract to receive book reviews. The provider of the reviews exposes the data in Amazon Simple Storage Service (Amazon S3) buckets. Litware plans to manage Search Engine Optimization (SEO) for the authors. The SEO data will be streamed from a REST API.

Requirements. Version Control

Litware plans to implement a version control solution in Fabric that will use GitHub integration and follow the principle of least privilege.

---

## Requirements. Governance Requirements

To control data platform costs, the data platform must use only Fabric services and items. Additional Azure resources must NOT be provisioned.

## Requirements. Data Requirements

Litware identifies the following data requirements:

- Process the SEO data in near-real-time (NRT).
- Make the book reviews available in the lakehouse without making a copy of the data.
- When a new book cover image arrives in the Files folder, process the image as soon as possible.

You need to implement the solution for the book reviews.

Which should you do?

- A. Create a Dataflow Gen2 dataflow.
- B. Create a shortcut.
- C. Enable external data sharing.
- D. Create a data pipeline.

Answer: B

Explanation:

The requirement specifies that Litware plans to make the book reviews available in the lakehouse without making a copy of the data. In this case, creating a shortcut in Fabric is the most appropriate solution. A shortcut is a reference to the external data, and it allows Litware to access the book reviews stored in Amazon S3 without duplicating the data into the lakehouse.

7. You need to resolve the sales data issue. The solution must minimize the amount of data transferred.

What should you do?

- A. Split the dataflow into two dataflows.
- B. Configure scheduled refresh for the dataflow.
- C. Configure incremental refresh for the dataflow. Set Store rows from the past to 1 Month.
- D. Configure incremental refresh for the dataflow. Set Refresh rows from the past to 1 Year.
- E. Configure incremental refresh for the dataflow. Set Refresh rows from the past to 1 Month.

Answer: E

Explanation:

The sales data issue can be resolved by configuring incremental refresh for the dataflow. Incremental refresh allows for only the new or changed data to be processed, minimizing the amount of data transferred and improving performance.

The solution specifies that data older than one month never changes, so setting the refresh period to 1 Month is appropriate. This ensures that only the most recent month of data will be refreshed, reducing unnecessary data transfers.

## 8.HOTSPOT

You need to troubleshoot the ad-hoc query issue.

How should you complete the statement? To answer, select the appropriate options in the answer area. NOTE: Each correct selection is worth one point.

### Answer Area

```
SELECT last_run_start_time, last_run_command
```

FROM

queryinsights.exec_requests_history
queryinsights.exec_sessions_history
queryinsights.frequently_run_queries
queryinsights.long_running_queries

```
WHERE last_run_total_elapsed_time_ms > 7200000
```

AND

max_run_total_elapsed_time_ms > 7200000
median_total_elapsed_time_ms > 7200000
number_of_canceled_runs > 1
number_of_failed_runs > 1
number_of_runs > 1

Answer:

## Answer Area

```
SELECT last_run_start_time, last_run_command
```

```
FROM
```

queryinsights.exec_requests_history
queryinsights.exec_sessions_history
queryinsights.frequently_run_queries
queryinsights.long_running_queries

```
WHERE last_run_total_elapsed_time_ms > 7200000
```

```
AND
```

max_run_total_elapsed_time_ms > 7200000
median_total_elapsed_time_ms > 7200000
number_of_canceled_runs > 1
number_of_failed_runs > 1
number_of_runs > 1

Explanation:

SELECT last\_run\_start\_time, last\_run\_command: These fields will help identify the execution details of the long-running queries.

FROM queryinsights.long\_running\_queries: The correct solution is to check the long-running queries using the queryinsights.long\_running\_queries view, which provides insights into queries that take longer than expected to execute.

WHERE last\_run\_total\_elapsed\_time\_ms > 7200000: This condition filters queries that took more than 2 hours to complete (7200000 milliseconds), which is relevant to the issue described.

AND number\_of\_failed\_runs > 1: This condition is key for identifying queries that have failed more than once, helping to isolate the problematic queries that cause failures and need attention.

Topic 3, Misc. Questions Set

---

9. You have a Fabric workspace.

You have semi-structured data.

You need to read the data by using T-SQL, KQL, and Apache Spark. The data will only be written by using Spark.

What should you use to store the data?

- A. a lakehouse
- B. an eventhouse
- C. a datamart
- D. a warehouse

Answer: A

Explanation:

A lakehouse is the best option for storing semi-structured data when you need to read it using T-SQL, KQL, and Apache Spark. A lakehouse combines the flexibility of a data lake (which can handle semi-structured and unstructured data) with the performance features of a data warehouse. It allows data to be written using Apache Spark and can be queried using different technologies such as T-SQL (for SQL-based querying), KQL (Kusto Query Language for querying), and Apache Spark (for distributed processing). This solution is ideal when dealing with semi-structured data and requiring a versatile querying approach.

10. You have a Fabric workspace that contains a warehouse named Warehouse1.

You have an on-premises Microsoft SQL Server database named Database1 that is accessed by using an on-premises data gateway.

You need to copy data from Database1 to Warehouse1.

Which item should you use?

- A. a Dataflow Gen1 dataflow
- B. a data pipeline
- C. a KQL queryset
- D. a notebook

Answer: B

Explanation:

To copy data from an on-premises Microsoft SQL Server database (Database1) to a warehouse (Warehouse1) in Microsoft Fabric, the best option is to use a data pipeline. A data pipeline in Fabric allows for the orchestration of data movement, from source to destination, using connectors, transformations, and scheduled workflows. Since the data is being transferred from an on-premises database and requires the use of a data gateway, a data pipeline provides the appropriate framework to facilitate this data movement efficiently and reliably.

11. You have a Fabric workspace that contains a warehouse named Warehouse1.

You have an on-premises Microsoft SQL Server database named Database1 that is

---

accessed by using an on-premises data gateway.

You need to copy data from Database1 to Warehouse1.

Which item should you use?

- A. an Apache Spark job definition
- B. a data pipeline
- C. a Dataflow Gen1 dataflow
- D. an eventstream

Answer: B

Explanation:

To copy data from an on-premises Microsoft SQL Server database (Database1) to a warehouse (Warehouse1) in Fabric, a data pipeline is the most appropriate tool. A data pipeline in Fabric is designed to move data between various data sources and destinations, including on-premises databases like SQL Server, and cloud-based storage like Fabric warehouses. The data pipeline can handle the connection through an on-premises data gateway, which is required to access on-premises data. This solution facilitates the orchestration of data movement and transformations if needed.

12. You have a Fabric F32 capacity that contains a workspace. The workspace contains a warehouse named DW1 that is modelled by using MD5 hash surrogate keys.

DW1 contains a single fact table that has grown from 200 million rows to 500 million rows during the past year.

You have Microsoft Power BI reports that are based on Direct Lake. The reports show year-over-year values.

Users report that the performance of some of the reports has degraded over time and some visuals show errors.

You need to resolve the performance issues.

The solution must meet the following requirements:

Provide the best query performance.

Minimize operational costs.

Which should you do?

- A. Change the MD5 hash to SHA256.
- B. Increase the capacity.
- C. Enable V-Order
- D. Modify the surrogate keys to use a different data type.
- E. Create views.

Answer: D

Explanation:

In this case, the key issue causing performance degradation likely stems from the use of MD5 hash surrogate keys. MD5 hashes are 128-bit values, which can be inefficient for large datasets like the 500 million rows in your fact table. Using a more efficient data type for surrogate keys (such as integer or bigint) would reduce the storage and



processing overhead, leading to better query performance. This approach will improve performance while minimizing operational costs because it reduces the complexity of querying and indexing, as smaller data types are generally faster and more efficient to process.

### 13.HOTSPOT

You have a Fabric workspace that contains a warehouse named DW1. DW1 contains the following tables and columns.

Table name	Column name	Description
SalesOrderDetail	ProductID	Contains the product ID of the ordered product
SalesOrderDetail	ModifiedDate	Contains the date of an order
SalesOrderDetail	OrderQty	Contains the order quantity
Product	ProductID	Contains the unique ID of a product
Product	Name	Contains a product name

You need to create an output that presents the summarized values of all the order quantities by year and product. The results must include a summary of the order quantities at the year level for all the products.

How should you complete the code? To answer, select the appropriate options in the answer area. NOTE: Each correct selection is worth one point.

Preguntas de prueba Microsoft DP-900 Obtén el examen en el ex

## Answer Area

(SO.ModifiedDate) AS OrderDate

SELECT CAST  
SELECT CONVERT  
SELECT YEAR

,P.Name AS ProductName

,SUM(SO.OrderQty) AS OrderQty

FROM [dbo].[SalesOrderDetail] SO

INNER JOIN [dbo].[Product] P

ON P.ProductID = SO.ProductID

GROUP BY

CUBE(YEAR(SO.ModifiedDate), P.Name)
GROUPING SETS ((YEAR(SO.ModifiedDate), P.Name), (YEAR(SO.ModifiedDate)))
ROLLUP(YEAR(SO.ModifiedDate), P.Name)
YEAR(SO.ModifiedDate), P.Name

ORDER BY OrderDate

Answer:

Preguntas de prueba Microsoft DP-71

## Answer Area

(SO.ModifiedDate) AS OrderDate

SELECT CAST

SELECT CONVERT

SELECT YEAR

,P.Name AS ProductName

,SUM(SO.OrderQty) AS OrderQty

FROM [dbo].[SalesOrderDetail] SO

INNER JOIN [dbo].[Product] P

ON P.ProductID = SO.ProductID

GROUP BY

CUBE(YEAR(SO.ModifiedDate), P.Name)

GROUPING SETS ((YEAR(SO.ModifiedDate), P.Name), (YEAR(SO.ModifiedDate)))

ROLLUP(YEAR(SO.ModifiedDate), P.Name)

YEAR(SO.ModifiedDate), P.Name

ORDER BY OrderDate

14. You have a Fabric workspace that contains a lakehouse named Lakehouse1. Data is ingested into Lakehouse1 as one flat table. The table contains the following columns.

Name	Description
TransactionID	Contains a unique ID for each transaction
Date	Contains the date of a transaction
ProductID	Contains a unique ID for each product
ProductColor	Contains a descriptive attribute that describes the color of each product
ProductName	Contains a unique name for each product
SalesAmount	Contains the sales amount of a transaction

You plan to load the data into a dimensional model and implement a star schema.

---

From the original flat table, you create two tables named FactSales and DimProduct. You will track changes in DimProduct.

You need to prepare the data.

Which three columns should you include in the DimProduct table? Each correct answer presents part of the solution. NOTE: Each correct selection is worth one point.

- A. Date
- B. ProductName
- C. ProductColor
- D. TransactionID
- E. SalesAmount
- F. ProductID

Answer: B, C, F

Explanation:

In a star schema, the DimProduct table serves as a dimension table that contains descriptive attributes about products. It will provide context for the FactSales table, which contains transactional data. The following columns should be included in the DimProduct table:

**ProductName:** The ProductName is an important descriptive attribute of the product, which is needed for analysis and reporting in a dimensional model.

**ProductColor:** ProductColor is another descriptive attribute of the product. In a star schema, it makes sense to include attributes like color in the dimension table to help categorize products in the analysis.

**ProductID:** ProductID is the primary key for the DimProduct table, which will be used to join the FactSales table to the product dimension. It's essential for uniquely identifying each product in the model.

15. You have a Fabric workspace named Workspace1 that contains a notebook named Notebook1.

In Workspace1, you create a new notebook named Notebook2.

You need to ensure that you can attach Notebook2 to the same Apache Spark session as Notebook1.

What should you do?

- A. Enable high concurrency for notebooks.
- B. Enable dynamic allocation for the Spark pool.
- C. Change the runtime version.
- D. Increase the number of executors.

Answer: A

Explanation:

To ensure that Notebook2 can attach to the same Apache Spark session as Notebook1, you need to enable high concurrency for notebooks. High concurrency allows multiple notebooks to share a Spark session, enabling them to run within the same Spark context and thus share resources like cached data, session state, and

---

compute capabilities. This is particularly useful when you need notebooks to run in sequence or together while leveraging shared resources.

16. You have a Fabric workspace named Workspace1 that contains a lakehouse named Lakehouse1.

Lakehouse1 contains the following tables:

- Orders
- Customer
- Employee

The Employee table contains Personally Identifiable Information (PII).

A data engineer is building a workflow that requires writing data to the Customer table, however, the user does NOT have the elevated permissions required to view the contents of the Employee table. You need to ensure that the data engineer can write data to the Customer table without reading data from the Employee table.

Which three actions should you perform? Each correct answer presents part of the solution. NOTE: Each correct selection is worth one point.

- A. Share Lakehouse1 with the data engineer.
- B. Assign the data engineer the Contributor role for Workspace2.
- C. Assign the data engineer the Viewer role for Workspace2.
- D. Assign the data engineer the Contributor role for Workspace1.
- E. Migrate the Employee table from Lakehouse1 to Lakehouse2.
- F. Create a new workspace named Workspace2 that contains a new lakehouse named Lakehouse2.
- G. Assign the data engineer the Viewer role for Workspace1.

Answer: A, D, E

Explanation:

To meet the requirements of ensuring that the data engineer can write data to the Customer table

without reading data from the Employee table (which contains Personally Identifiable Information, or

PII), you can implement the following steps:

Share Lakehouse1 with the data engineer.

By sharing Lakehouse1 with the data engineer, you provide the necessary access to the data within the lakehouse. However, this access should be controlled through roles and permissions, which will allow writing to the Customer table but prevent reading from the Employee table.

Assign the data engineer the Contributor role for Workspace1.

Assigning the Contributor role for Workspace1 grants the data engineer the ability to perform actions such as writing to tables (e.g., the Customer table) within the workspace. This role typically allows users to modify and manage data without necessarily granting them access to view all data (e.g., PII data in the Employee table).

---

Migrate the Employee table from Lakehouse1 to Lakehouse2.

To prevent the data engineer from accessing the Employee table (which contains PII), you can migrate the Employee table to a separate lakehouse (Lakehouse2) or workspace (Workspace2). This separation of sensitive data ensures that the data engineer's access is restricted to the Customer table in Lakehouse1, while the Employee table can be managed separately and protected under different access controls.

17. You have a Fabric warehouse named DW1. DW1 contains a table that stores sales data and is used by multiple sales representatives.

You plan to implement row-level security (RLS).

You need to ensure that the sales representatives can see only their respective data.

Which warehouse object do you require to implement RLS?

- A. STORED PROCEDURE
- B. CONSTRAINT
- C. SCHEMA
- D. FUNCTION

Answer: D

Explanation:

To implement Row-Level Security (RLS) in a Fabric warehouse, you need to use a function that defines the security logic for filtering the rows of data based on the user's identity or role. This function can be used in conjunction with a security policy to control access to specific rows in a table.

In the case of sales representatives, the function would define the filtering criteria (e.g., based on a column such as SalesRepID or SalesRepName), ensuring that each representative can only see their respective data.

## 18. HOTSPOT

You have a Fabric workspace named Workspace1\_DEV that contains the following items:

- 10 reports
- Four notebooks
- Three lakehouses
- Two data pipelines
- Two Dataflow Gen1 dataflows
- Three Dataflow Gen2 dataflows
- Five semantic models that each has a scheduled refresh policy

You create a deployment pipeline named Pipeline1 to move items from Workspace1\_DEV to a new workspace named Workspace1\_TEST.

You deploy all the items from Workspace1\_DEV to Workspace1\_TEST.

For each of the following statements, select Yes if the statement is true. Otherwise,

select No. NOTE: Each correct selection is worth one point.

### Answer Area

Statements	Yes	No
Data from the semantic models will be deployed to the target stage.	<input type="radio"/>	<input type="radio"/>
The Dataflow Gen1 dataflows will be deployed to the target stage.	<input type="radio"/>	<input type="radio"/>
The scheduled refresh policies will be deployed to the target stage.	<input type="radio"/>	<input type="radio"/>

Answer:

### Answer Area

Statements	Yes	No
Data from the semantic models will be deployed to the target stage.	<input type="radio"/>	<input checked="" type="radio"/>
The Dataflow Gen1 dataflows will be deployed to the target stage.	<input checked="" type="radio"/>	<input type="radio"/>
The scheduled refresh policies will be deployed to the target stage.	<input type="radio"/>	<input checked="" type="radio"/>

19. You have a Fabric deployment pipeline that uses three workspaces named Dev, Test, and Prod.

You need to deploy an eventhouse as part of the deployment process.

What should you use to add the eventhouse to the deployment process?

- A. GitHub Actions
- B. a deployment pipeline
- C. an Azure DevOps pipeline

Answer: B

Explanation:

A deployment pipeline in Fabric is designed to automate the process of deploying assets (such as reports, datasets, eventhouses, and other objects) between environments like Dev, Test, and Prod. Since you need to deploy an eventhouse as

---

part of the deployment process, a deployment pipeline is the appropriate tool to move this asset through the different stages of your environment.

20. You have a Fabric workspace named Workspace1 that contains a warehouse named Warehouse1.

You plan to deploy Warehouse1 to a new workspace named Workspace2.

As part of the deployment process, you need to verify whether Warehouse1 contains invalid references. The solution must minimize development effort.

What should you use?

- A. a database project
- B. a deployment pipeline
- C. a Python script
- D. a T-SQL script

Answer: C

Explanation:

A deployment pipeline in Fabric allows you to deploy assets like warehouses, datasets, and reports between different workspaces (such as from Workspace1 to Workspace2). One of the key features of a deployment pipeline is the ability to check for invalid references before deployment. This can help identify issues with assets, such as broken links or dependencies, ensuring the deployment is successful without introducing errors. This is the most efficient way to verify references and manage the deployment with minimal development effort.

21. You have a Fabric workspace that contains a Real-Time Intelligence solution and an eventhouse.

Users report that from OneLake file explorer, they cannot see the data from the eventhouse.

You enable OneLake availability for the eventhouse.

What will be copied to OneLake?

- A. only data added to new databases that are added to the eventhouse
- B. only the existing data in the eventhouse
- C. no data
- D. both new data and existing data in the eventhouse
- E. only new data added to the eventhouse

Answer: D

Explanation:

When you enable OneLake availability for an eventhouse, both new and existing data in the eventhouse will be copied to OneLake. This feature ensures that data, whether newly ingested or already present, becomes available for access through OneLake, making it easier for users to interact with and explore the data directly from OneLake file explorer.



22. You have a Fabric workspace named Workspace1. You plan to integrate Workspace1 with Azure DevOps. You will use a Fabric deployment pipeline named deployPipeline1 to deploy items from Workspace1 to higher environment workspaces as part of a medallion architecture. You will run deployPipeline1 by using an API call from an Azure DevOps pipeline. You need to configure API authentication between Azure DevOps and Fabric. Which type of authentication should you use?

- A. service principal
- B. Microsoft Entra username and password
- C. managed private endpoint
- D. workspace identity

Answer: A

Explanation:

When integrating Azure DevOps with Fabric (Workspace1) using a service principal is the recommended authentication method. A service principal provides a way for applications (such as an Azure DevOps pipeline) to authenticate and interact with resources securely. It allows Azure DevOps to authenticate API calls to Fabric without requiring direct user credentials. This method is ideal for automating tasks such as deploying items through a Fabric deployment pipeline.

23. You have a Google Cloud Storage (GCS) container named storage1 that contains the files shown in the following table.

Name	Size
ProductFile.parquet	8 MB
StoreFile.json	500 MB
TripsFile.csv	99 MB

You have a Fabric workspace named Workspace1 that has the cache for shortcuts enabled. Workspace1 contains a lakehouse named Lakehouse1. Lakehouse1 has the shortcuts shown in the following table.

Name	Source	Last accessed
Products	ProductFile	12 hours ago
Stores	StoreFile	4 hours ago
Trips	TripsFile	48 hours ago

You need to read data from all the shortcuts.

Which shortcuts will retrieve data from the cache?

- A. Stores only
- B. Products only
- C. Stores and Products only
- D. Products, Stores, and Trips
- E. Trips only
- F. Products and Trips only

Answer: C

Explanation:

When reading data from shortcuts in Fabric (in this case, from a lakehouse like Lakehouse1), the cache for shortcuts helps by storing the data locally for quick access. The last accessed timestamp and the cache expiration rules determine whether data is fetched from the cache or from the source (Google Cloud Storage, in this case).

Products: The ProductFile.parquet was last accessed 12 hours ago. Since the cache has data available for up to 12 hours, it is likely that this data will be retrieved from the cache, as it hasn't been too long since it was last accessed.

Stores: The StoreFile.json was last accessed 4 hours ago, which is within the cache retention period.

Therefore, this data will also be retrieved from the cache.

Trips: The TripsFile.csv was last accessed 48 hours ago. Given that it's outside the typical caching window (assuming the cache has a maximum retention period of around 24 hours), it would not be retrieved from the cache. Instead, it will likely require a fresh read from the source.

24. You have a Fabric workspace named Workspace1 that contains an Apache Spark job definition named Job1.

You have an Azure SQL database named Source1 that has public internet access disabled.

You need to ensure that Job1 can access the data in Source1.

What should you create?

- A. an on-premises data gateway

- B. a managed private endpoint
- C. an integration runtime
- D. a data management gateway

Answer: B

Explanation:

To allow Job1 in Workspace1 to access an Azure SQL database (Source1) with public internet access disabled, you need to create a managed private endpoint. A managed private endpoint is a secure, private connection that enables services like Fabric (or other Azure services) to access resources such as databases, storage accounts, or other services within a virtual network (VNet) without requiring public internet access. This approach maintains the security and integrity of your data while enabling access to the Azure SQL database.

25. You have an Azure Data Lake Storage Gen2 account named storage1 and an Amazon S3 bucket named storage2.

You have the Delta Parquet files shown in the following table.

Name	Stored in	Size	Description
ProductFile	storage1	50 MB	Contains a list of products and their details
TripsFile	storage2	2 GB	Contains one month's worth of taxi trip data
StoreFile	storage2	25 MB	Contains a list of stores and their addresses

You have a Fabric workspace named Workspace1 that has the cache for shortcuts enabled.

Workspace1 contains a lakehouse named Lakehouse1.

Lakehouse1 has the following shortcuts:

- A shortcut to ProductFile aliased as Products
- A shortcut to StoreFile aliased as Stores
- A shortcut to TripsFile aliased as Trips

The data from which shortcuts will be retrieved from the cache?

- A. Trips and Stores only
- B. Products and Store only
- C. Stores only
- D. Products only
- E. Products, Stores, and Trips

Answer: B

Explanation:

When the cache for shortcuts is enabled in Fabric, the data retrieval is governed by the caching behavior, which generally retains data for a specific period after it was last accessed. The data from the shortcuts will be retrieved from the cache if the data is stored in locations that support caching.

---

Here's a breakdown based on the data's location:

**Products:** The ProductFile is stored in Azure Data Lake Storage Gen2 (storage1). Since Azure Data Lake is a supported storage system in Fabric and the file is relatively small (50 MB), this data is most likely cached and can be retrieved from the cache.

**Stores:** The StoreFile is stored in Amazon S3 (storage2), and even though it is stored in a different cloud provider, Fabric can cache data from Amazon S3 if caching is enabled. This data (25 MB) is likely cached and retrievable.

**Trips:** The TripsFile is stored in Amazon S3 (storage2) and is significantly larger (2 GB) compared to the other files. While Fabric can cache data from Amazon S3, the larger size of the file (2 GB) may exceed typical cache sizes or retention windows, causing this file to likely be retrieved directly from the source instead of the cache.

**Preguntas de prueba Microsoft DP-700 - Obtén éxito en el examen DP-700 sin estrés**

# KILLTEST

---

20% de descuento en todos los productos

**Obtenga la versión  
completa de las  
preguntas del examen**

[DP-700](#)

**Killtest.es**

